Mon. Not. R. Astron. Soc. **000**, 1–13 (2003) Printed 23 January 2003 (MN LATEX style file v1.4)

Image Simulation with Shapelets

Richard J. Massey[1⋆], Alexandre R. Refregier[2,1,3], Christopher J. Conselice[3] & David J. Bacon[4]

[1] *Institute of Astronomy, Madingley Road, Cambridge CB3 OHA, UK*
[2] *Service d'Astrophysique, Bât. 709, CEA Saclay, F-91191 Gif sur Yvette, France*
[3] *California Institute of Technology, 1201 E. California Blvd., Pasadena CA 91125, USA*
[4] *Institute for Astronomy, Blackford Hill, Edinburgh EH9 3HJ, UK*

Accepted —. Received —; in original form —.

ABSTRACT

We present a method to manufacture simulated deep sky images, with realistic galaxy morphologies and telescope characteristics. For this purpose, we first use the shapelets formalism (Refregier 2003) to parametrize the shapes of all galaxies in the Hubble Deep Field. We consider the distribution of real galaxy morphologies in shapelet parameter space, then resample this distribution to generate new galaxies. The simulated objects include realistic spiral arms, bars, discs, arbitrary radial profiles and even dust lanes or knots. We apply standard morphology diagnostics to demonstrate that our artificial images closely mimic real data in terms of galaxy size, concentration and asymmetry statistics, *etc.* Observational effects, including Point-Spread Function, noise, pixellisation, and astrometric distortions are also modelled. Sample images are made available on the world wide web. These simulations are useful to develop and test precision image analysis techniques, including photometry, astrometry, and shape measurement. The sensitivity of future telescopes and surveys can thus be assessed for applications including supernova searches, microlensing, proper motions, and weak lensing.

Key words: galaxies: fundamental parameters – galaxies: statistics – methods: statistical – gravitational lensing.

1 INTRODUCTION

As astronomical surveys are growing in size and scope, so image analysis methods are increasing in complexity and accuracy. In order to calibrate these image analysis methods, it is essential to have a large sample of images which contain objects with known properties. Since real data is subject to the uncertainties of observational effects, several packages have been developed to generate artificial, simulated images (*e.g.* `Skymaker` (see Erben *et al.* 2001) or `artdata` in `IRAF` (Tody 1993)). The accuracy of the image analysis methods can then be assessed by comparing their output to the input image properties specified before the addition of complications like noise, convolution with a Point-Spread Function (PSF) and other observational systematics.

The simulation packages currently available are particularly valuable for creating simulations of deep ground-based imaging data. However, they limit themselves to a representation of galaxies as parametric forms such as symmetric de Vaucouleur or exponential profiles. Deep space-based images, on the other hand, contain many irregular or asymmetrical galaxies with complex resolved features such as spiral arms, mergers and dust lanes, *etc.* One possibility for simulating space images, utilised by Bouwens, Broadhurst & Silk (1998) is to reuse real, well-resolved galaxies from the Hubble Deep Fields (HDFs; Williams *et al.* 1996; Williams *et al.* 1998) in simulated images. However, this restricts us to using the same real galaxies many times in a large simulation. Although the HDFs are indeed very deep (I_{F814W}=27.60 at 10σ, Williams *et al.* 1996), they have a small angular size (∼6 square arcminutes each). Even if we were to use larger surveys such as the Groth strip (Groth *et al.* 1994) or the Medium Deep Survey (Ratnatunga, Griffiths & Ostrander 1999) as the source of our real galaxies, we would still face the difficulty of using particular real galaxies many times in our simulation. This method also faces the difficulty that, if we were to lay down the faintest real galaxies, their morphological properties would have been significantly contaminated with background noise, and would therefore not fairly represent the faint galaxy population.

Here, we present a method for simulating deep images which generates objects that are genuinely new, yet replicate the morphological properties of galaxies in the HDF at all depths. This procedure has the advantage of allowing us

⋆ E-mail: rjm@ast.cam.ac.uk

to simulate arbitrarily large, deep surveys with no repetition of galaxy shapes. It also allows us to know accurately the intrinsic properties for each galaxy, before adding telescope-specific noise properties (*e.g.* Poisson photon noise) and systematic errors (*e.g.* PSF smearing).

Our method is to decompose all objects in the HDFs into shapelet parametrizations, following the formalism of Refregier (2003, hereafter paper I) and Refregier & Bacon (2003; hereafter paper II). Using just a few coefficients, these can completely quantify the shape properties of all galaxies, including spiral arms, bars and arbitrary radial profiles. We model the distribution of these shapelet coefficients, and then draw from this probability distribution new sets of shapelet coefficients, representing new galaxies. In particular, we take into account the covariance between shapelet coefficients so that, for example, shapes depend upon magnitude (*e.g.* faint galaxies appear more irregular than bright ones). In this method, we therefore do not input any model of physical morphology or evolution. Rather, we exclusively use the measured statistics of shapelet coefficient distributions from a real galaxy sample. The new galaxy images can then be analytically convolved with any PSF and given an appropriate amount of noise for any exposure time down to the depth of the HDF.

These simulations have several significant applications. We can use them to calibrate the effectiveness of image analysis and detection methods such as `SExtractor` (Bertin & Arnouts 1996), `imcat` (Kaiser, Squire & Broadhurst 1995), `GIM2D` (Simard 1998), `GALFIT` (Peng *et al.* 2002) and wavelet routines (*e.g.* Meyer 1993). By examining the noise levels on shape measurement at various signal-to-noise levels of galaxy detection, we can also estimate the accuracy of future experiments requiring accurate shape measurement (*e.g.* cosmic shear surveys with space-based telescopes; *c.f.* Massey *et al.* 2003b).

The paper is organized as follows. In §2 we give a brief overview of the shapelet formalism, and describe how the HDF galaxies are decomposed into shapelet parameters. We then discuss in §3 the means by which we find a smooth probability distribution in shapelet parameter space from which to generate new objects. We discuss examples of the resulting simulated images in §4.

We then demonstrate that the simulations do indeed have similar properties to the HDFs. For this purpose, we consider in §5 commonly used quantifiers for galaxy morphology. These include the size-magnitude distribution and ellipticity distribution of galaxies using `SExtractor` (Bertin & Arnouts 1996), and good agreement for measures such as concentration, asymmetry and clumpiness indices (*e.g.* Bershady *et al.* 2000, Conselice *et al.* 2000a). We summarise our findings in §6. Sample images may be downloaded from `http://www.ast.cam.ac.uk/~rjm/shapelets`.

2 SHAPELET SOURCE CATALOGUE

In this section, we describe the detection and modelling of HDF galaxies using shapelets. This procedure creates a catalogue of galaxy morphologies, which we will use in later sections to generate simulated galaxies.

2.1 Source detection

Objects are initially detected using `SExtractor` (Bertin & Arnouts 1996) in the HDF images and the pixel weight maps outputted by `DRIZZLE` (Fruchter & Hook 2002). A convolution mask and detection parameters were adapted from those used by Williams *et al.* (1996). However, the S/N detection threshold, DETECT_THRESH, is lowered to 1.3. These settings recover faint galaxies and minimize incompleteness at the expense of many false-positive 'detections' of noise, which will be filtered out later (see §2.2). Stars with CLASS_STAR $> 97\%$ are immediately discarded. The image is then segmented into small 'postage stamp' regions around the remaining galaxies. The sizes of these regions are set to $(3 \times \mathrm{A\_IMAGE} + 5)$ pixels square, where A_IMAGE is a measure of the galaxy major axis provided by `SExtractor`. The area is slightly smaller than those illustrated in figure 1, and square even when the object is highly elliptical. This is compact enough to be computationally efficient, but large enough to ensure that the shapelet basis functions are completely contained within the postage stamp images.

This prescription also leaves a small border of background noise conveniently around the edge of each image. We use all border pixels that belong to no other object in the `SExtractor` catalogue, to locally renormalise the pixel weight map. As noticed by Williams *et al.* (1996), the inverse variance map outputted during the data reduction of the HDF systematically overestimates the noise by a factor of a few. This bias also varies as a function of position around the image. While `SExtractor` requires only relative weights between pixels, and is thus unaffected by this bias, we need to calibrate the absolute value of noise. Once this is correct, the shapelet model can be fit to capture the object's shape but ignore the noise.

2.2 Shapelet decomposition

Shapelets are a complete, orthonormal set of 2D basis functions. A linear combination of these functions can be used to model any image, similarly to Fourier or wavelet synthesis. The shapelet decomposition is particularly efficient for images localised in space, such as those of individual galaxies. The formalism was first introduced in paper I, and a related method has also been independently suggested by Bernstein & Jarvis (2002).

For the polar shapelet analysis, the surface brightness $f(\mathbf{x})$ of an object is expressed as

$$f(\mathbf{x}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_{nm} \phi_{nm}(\mathbf{x} - \mathbf{x}_c; \beta) \ , \qquad (1)$$

where β is a scale parameter and $\mathbf{x}_c$ is the position of the centre of the basis functions. The basis functions, ϕ_{nm}, expressed in their polar separable form, are given by

$$\phi_{n,m}(r,\theta) \equiv \frac{L_{n,m}\left(\frac{r}{\beta}\right)\ e^{-\frac{r^2}{2\beta^2}}\ e^{im\theta}}{\beta\sqrt{\pi\left(\frac{n+m}{2}\right)!\left(\frac{n-m}{2}\right)!}} \ , \qquad (2)$$

where $L_{n,m}(x)$ are the Laguerre polynomials (see *e.g.* Boas 1983). The index n describes the number of radial oscillations and the index m describes the order of rotational symmetry. The shapelet coefficients a_{nm} are Gaussian-weighted

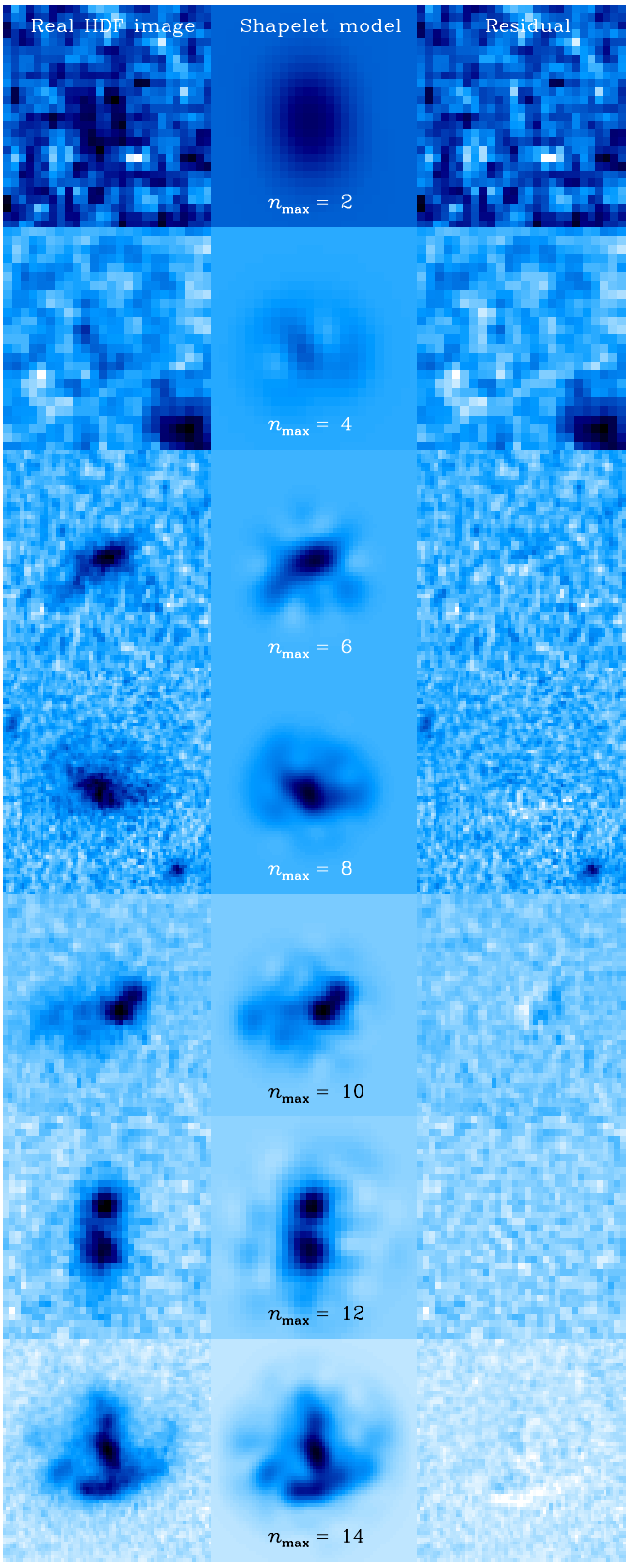

Figure 1. Shapelet modelling of a selection of HDF I-band galaxies. Higher S/N galaxies typically require more shapelet coefficients and we display a variety of source galaxies, noting the shapelet $n_{\max}$ required to reach a reconstruction with $\chi^2_r = 1 \pm 1\sigma$. In all cases, the first column shows the original HDF image; the middle column shows the shapelet model; the right column shows the residual. The image size and colour scale is different for each row.

multipole moments of the surface brightness, familiar in several branches of astronomy.

For reasonable choices of the centroid $\mathbf{x}_c$ and scale size β, the galaxy shape information is contained within only the first few shapelet coefficients. The series in equation (1) can then be truncated at some finite order $n_{\max}$. To implement this choice, we first define χ^2_r as the difference between the original and reconstructed image, renormalised with respect to the local noise level (see §2.1). Then we iterate values of $\mathbf{x}_c$ and β to find the shapelet parameters a_{nm} which achieve $\chi^2_r = 1 \pm 1\sigma$ using a minimum possible $n_{\max}$. All higher n shapelet coefficients are subsequently discarded. This algorithm outputs a catalogue of shapelet decompositions for ~ 500 objects per square arcminute in the HDFs. Our method will be described in detail in a future paper.

In figure 1 we present an initial indication that it successfully captures the shapes of galaxies. This displays a selection of HDF galaxies at various S/N levels, and their shapelet reconstructions (see also paper I, figures 3 and 4). Typical faint galaxies require an $n_{\max}$ of only 2, 3 or 4, while brighter objects require an increasing number of shapelet coefficients. The right-hand column of figure 1 shows the reconstruction residuals, which are consistent with noise even for irregular galaxy morphologies. We find the results superior to those of *e.g.* `GALFIT` (Peng *et al.* 2002), using traditional radial galaxy profiles alone.

The shapelets parametrization has several advantages. The truncation in $n_{\max}$ produces data compression by setting a minimum and maximum physical scale of interest. For example, the discarded high-n order coefficients contain a small amount of high spatial frequency information. Since we have ensured that $\chi^2_r \simeq 1$, the recovery is already acceptable and this is consistent with noise. Any artifact residuals are also at the level of the background noise. The compression factor for typical galaxy morphologies is comparable to or better than that possible using wavelets and can be as high as 50× (paper I). Furthermore, this is achieved through a parameter-independent truncation of a series. Shapelets avoids the requirement in `GALFIT` or `GIM2D` (Simard 1998) to specify in advance the number and type of profiles for each model, by using a complete basis set. A Karhunen-Loéve decomposition also requires a model of both the image and the noise. Instead, the shapelet parametrization is uniquely specified by a linear, one-to-one mapping from the image plane to the coefficients. Moreover, our orthonormal basis set has many convenient mathematical properties that are lost to methods using an overcomplete basis set such as `Pixon` (Piña & Puetter 1993). It is for these mathematical properties that we choose to model galaxies with shapelets, rather than anything else.

Shapelets are designed in particular to be convenient for many aspects of image manipulation. The shapelet basis functions ϕ_{nm} are also the eigenfunctions of the 2D Quantum Harmonic Oscillator. As a result, they are invariant under Fourier transform, up to a phase factor. This renders convolutions (*e.g.* with a PSF) easy and quick to perform. It also suggests a well-developed mathematical notation. Convolutions become a bra-ket matrix multiplication (see paper II). Translations, dilations and other linear transformations (such as the shears that we study in weak gravitational lensing) are described to first order by a few applications of

quantum mechanical $\hat{a}$ and $\hat{a}^\dagger$ ladder operators (see paper I).

The current implementation of the $\mathbf{x}_c$ and β iteration does not converge for all objects in the `SExtractor` catalogue of the HDFs. It fails for a total of ∼19% of the 'objects' detected by `SExtractor`. Of these, we find that ∼8% are false detections of noise, due to the low S/N detection threshold set for `SExtractor` (*c.f.* section §2.1). It also fails for 6% of the objects, which are close galaxy pairs, and an additional ∼5% more of the objects for miscellaneous reasons such as proximity to bright stars or the edge of the image. Note that the decompositions which fail due to close galaxy pairs are independent of magnitude and therefore do not cause a bias in the sampled galaxy population. Indeed, the slope of the number counts for galaxy pair members is within $\sim 1\sigma$ of that for all the galaxies in the HDF. The shapelet catalogue does conspicuously lack large objects ($\gtrsim 2.5''$) which may be a consequence of their increased likelihood of having near neighbours. The impact of these effects are small, given the fact that the HDF is dominated by small, faint objects.

The postage stamp images extracted around each HDF object in §2.1 are thus decomposed into shapelets. The resulting data vectors of shapelet coefficients encode the shape information about each HDF object. Collectively, they describe the overall morphology distribution of distant galaxies. This will be used in §3 to obtain a description of the galaxy morphology PDF.

2.3 Treatment of the PSF

During the modelling of galaxy shapes, we must account for the PSF of the WFPC2 camera which has smeared the HDF images. Since our objective here is to simulate only HST images, we do not correct for the PSF. It is naturally contained within the shapelet models of the galaxy images and unaltered. Our simulated images will be automatically smeared by an effective circularised WFPC2 PSF, arising from the random orientations of the galaxies.

However, for other applications it may be desirable to simulate observations from other telescopes such as the JWST (`http://www.stsci.edu/ngst/`) , SNAP (`http://snap.lbl.gov/`) or GAIA (`http://astro.esa.int/gaia/`). It would then be necessary to take account of their different instrumental properties. The ideal way to do this would be to deconvolve HDF galaxies from the WFPC2 PSF during the shapelet fitting procedure (see paper II), and then to reconvolve simulated galaxies with a new PSF at the end. Unfortunately, we have found this method difficult to implement in practice. The process of deconvolution naturally pushes information into high-n and m shapelet coefficients (see paper II). Although the ensuing galaxy reconstructions are still realistic, information about the overall galaxy morphology distribution is spread thinly around an increased number of coefficients. This distribution is no longer sufficiently well sampled by galaxies in the HDFs for the smoothing-and-resampling method presented in §3 to be effective.

An alternative solution exists to simulate images with a PSF of the same size or larger than that of HST. The WFPC2 PSF can be conveniently maintained throughout the simulations, and the images convolved again at the end with a second, 'difference' kernel. This kernel is intended to make up the difference between the original PSF of WFPC2 and that of the new instrument. It can be obtained by deconvolving the WFPC2 PSF from the new PSF, an operation performed easily in shapelet space (see paper II). An example of this method can be seen in Massey *et al.* 2003b.

3 SHAPELET PARAMETER SPACE

3.1 The multi-dimensional Hubble Tuning Fork

A sample of galaxy morphologies can be thought of as a distribution of points in a multi-dimensional shape parameter space. The axes in this space might represent size, magnitude, position angle (P.A.) and so on. Each point corresponds to a particular real galaxy, and various correlations may emerge between variables. For example, the classic Hubble tuning-fork diagram (Hubble 1926, Sandage 1961, de Vaucouleurs 1959) relates the object ellipticity, the bulge/disc ratio, and the extent to which the spiral arms are unwound. `GIM2D` (Simard 1998) and `GALFIT` (Peng 2002) software use axes representing the amount of an exponential disc or de Vaucouleurs/Sérsic profile (de Vaucouleurs 1959, Sérsic 1968) required to fit the galaxy's radial profile.

In this work, we instead choose as axes the shapelet coefficients that were measured for galaxies in §2. In §3.2, we describe this proposed 'shapelet parameter space'. The finite number of points in this space corresponding to real HDF galaxies represent valid shapes for galaxies. In §3.3 we argue that their observed distribution is a sampling from an underlying Probability Density Function (PDF) of galaxy morphology. We justify, from a morphological perspective, our procedure to recover the underlying PDF by smoothing the parameter space. In §4, we show how the PDF can then be Monte-Carlo sampled to produce simulated galaxies.

3.2 The multi-dimensional Shapelet Tuning Fork

The polar shapelet formalism is convenient because an object's orientation is controlled by the phase of the a_{22} coefficient, and its chirality (handedness) by the relative phase of the a_{42} coefficient. The first can be factored easily out of the parametrization, so that the ellipticity of all objects becomes horizontal. The image can be flipped, if necessary, so that the sign of the second is positive and the outer isophotes of all objects twist in the same anti-clockwise sense. Correlations between remaining shapelet coefficients a_{nm} are maintained in order to preserve the morphology of each real galaxy. Any two well-sampled objects which are identical other than their P.A. will then decompose into identical shapelet coefficients, at the same location in shapelet parameter space. Assuming that P.A. is intrinsically random, this procedure compresses the morphology parameter space without loss of information, enabling it to be more densely sampled by the same finite number of HDF galaxies. Simulated galaxies will be randomly re-rotated and flipped as they are created.

Slices through the compressed shapelet parameter space are shown in figure 2. Notice that there are correlations evident in the parameters, which correspond to the construction of the familiar shapes of galaxies. For example, the scatter of ellipticity values widens for faint galaxies in the middle-left plot. In the bottom-right plot, deviations from

the diagonal show twisting isophotes that can grow into spiral arms. Above all, notice that some regions of parameter space are empty. A random set of shapelet coefficients will not produce a realistic galaxy shape: in particular there is no positive definite constraint imposed upon an image in the shapelet formalism. That our simulated galaxies do not contain negative holes will be a basic test in section §4.2.

Two other parameters are required, since real galaxy morphologies clearly vary as a function of size and magnitude (*e.g.* figure 5). Storing the shapelet scale factor β (see §2.2) allows large HDF galaxies to occupy different regions of parameter space than small ones. This is also the case for an object's magnitude. Since shapelet coefficients (including a_{00}) scale as the flux, we choose to implement this by dividing all coefficients a_{nm} by a_{00}. This removes explicit magnitude dependence in these quantities and coincidentally ensures a convenient version of adaptive smoothing at a later stage (see §4.1). Since now $a_{00} = 1$ for every object, this degenerate parameter is removed. Size and magnitude are treated as any other axes of the parameter space (see figure 2).

Note that *any* orthogonal transformation of the shapelet basis functions would maintain their useful properties of completeness, orthogonality and Fourier transform invariance. For instance, the Cartesian version of shapelets can be used instead (see paper I), but without the convenient factoring out of the object's orientation. Using principal components analysis (PCA; *e.g.* Francis & Wills 1999), it is possible to calculate the optimal linear combination of shapelet coefficients to quantitatively describe galaxy morphology with fewest numbers. However, both elliptical and spiral galaxy shapes are already simple to manufacture by specifying only a few polar shapelet coefficients. We therefore avoid the extra complication of PCA in this paper. However, the principle components of galaxy morphology are interesting in their own right and are being studied elsewhere.

3.3 Recovery of a smooth underlying PDF

In the top-left panel of figure 2, we show a slice through the parameter space of galaxy morphologies, populated by δ-functions representing real, observed shapes. Unlike a distribution parametrized simply by bulge/disc ratios and disc inclination angles, it is not obvious *a priori* that an underlying, smooth PDF should exist for shapelets. Nonetheless, the compact shapelet representation of astronomical objects suggests that this ought to be the case.

Figure 2 demonstrates that the parameter space is indeed smooth on large scales, in those places where it is well sampled. We assume that some other regions are poorly sampled merely because of the finite number of galaxies in the HDF. We acknowledge that there are also voids in the parameter space, which do not correspond to realistic galaxy shapes and we do not allow heavy smoothing which would significantly encroach upon these voids. However, limited perturbations around HDF galaxies will indeed recover realistic morphologies.

Without an explicitly physical model of galaxy morphology and evolution built in to shapelets, it is the final results that must provide the ultimate verification of our statistical method. In §4.2 and figure 3 below, we show that it is possible to find a smoothing length for the PDF that recovers realistic-looking objects. In §5 we demonstrate quantitatively that their global properties are realistic by comparing real and simulated populations of galaxies via morphology diagnostics commonly used on deep images.

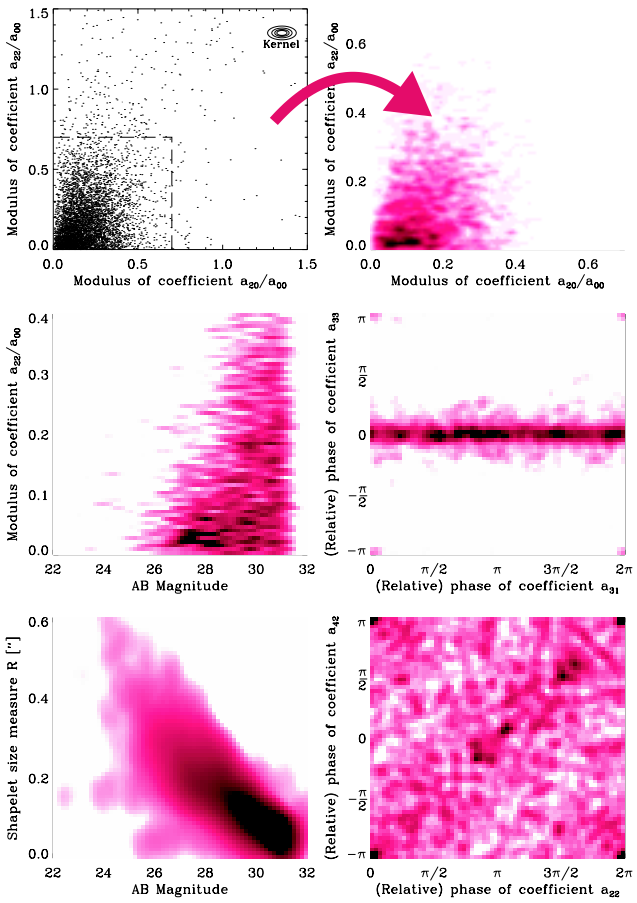

Figure 2. Phase space correlations and smoothing in the shapelet parameter space. The top left panel displays the position of measured HDF galaxies along two axes of shapelet space; the top right panel shows the probability distribution produced by smoothing this distribution. The other left panels display further projections of the PDF onto shapelet coefficient, size and magnitude axes, while the remaining right panels display phase correlations between shapelet coefficients. The colour scale is logarithmic in the bottom left panel.

It is then an easy step to resample the recovered underlying PDF ourselves, and thus to synthesise our own population. Monte Carlo techniques can be used to generate unlimited numbers of new galaxies, all different yet realistic, to fill any amount of sky area in a simulated imaging survey.

4 IMAGE GENERATION

4.1 Multivariate kernel smoothing method

Many statistical approaches have been devised to smooth discrete samplings of a multivariate PDF and draw new simulated populations. Our main constraint in selecting one of these methods is the very high dimensionality of our data set. The median $n_{\max}$ required for objects in the HDF is

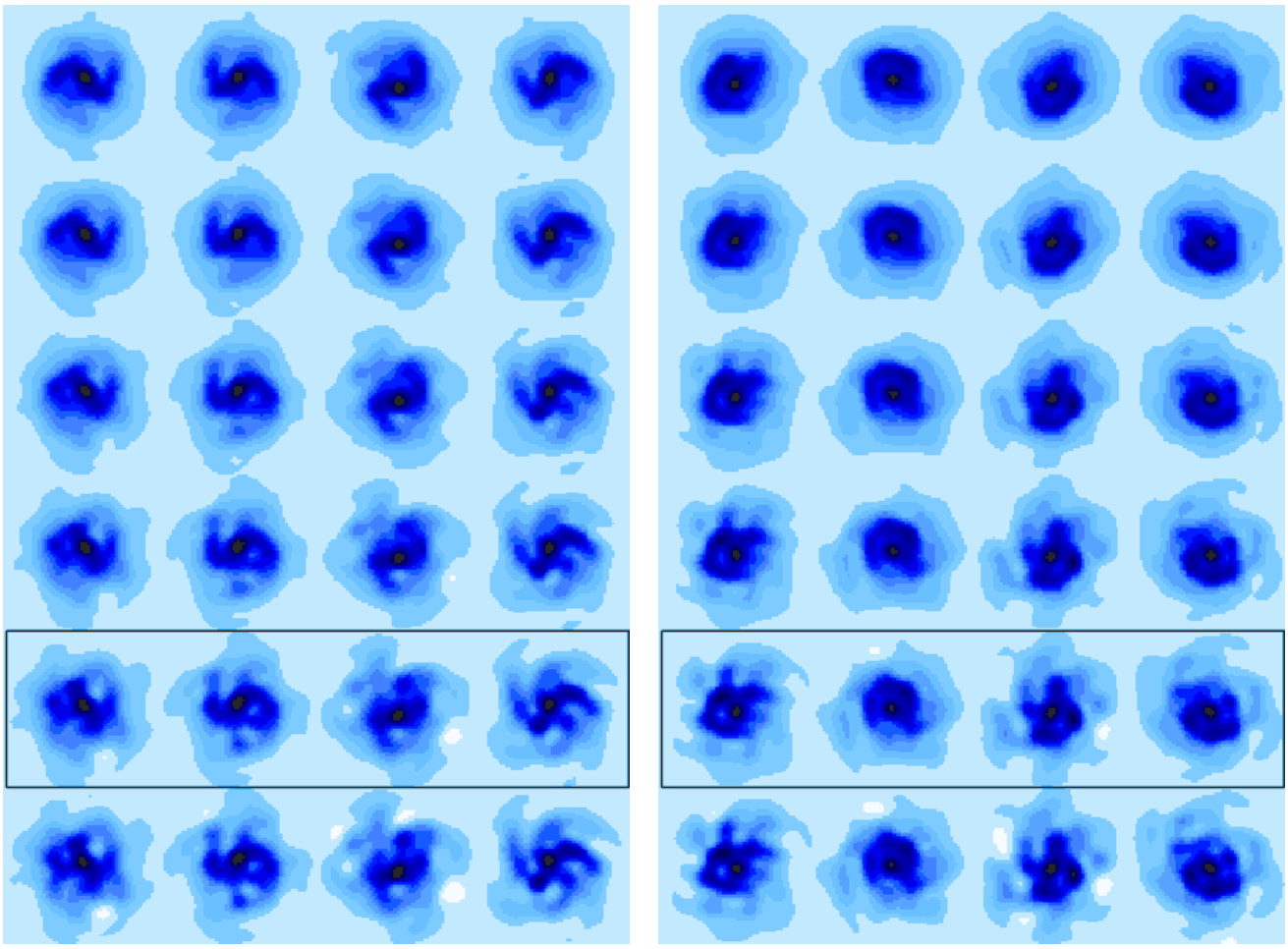

Figure 3. The effect of perturbing galaxy morphologies in shapelet space. Each image in the top row shows a real HDF galaxy, rotated by random angles. Its shapelet coefficients are incrementally perturbed in successive rows, although its overall flux is kept constant for the purpose of this plot. A degree of perturbation corresponding to our choice of the smoothing length λ_i is shown inside the box: these images represent typical simulated galaxies. Perturbations larger than λ_i produce objects which contain significantly negative pixel values. The left panel depicts a spiral galaxy; the right panel a more typical irregular form. The colour scale is logarithmic.

4. However, even with the efficient data compression that shapelets can afford, models of the highest S/N galaxies use values for $n_{\max}$ as high as 15. Adding object size and magnitude, this corresponds to 137 total coefficients, and is the maximum number of dimensions required.

To smooth and resample this dataset, we have chosen the Kernel smoothing method which is reviewed by Silverman (1986). Kernel smoothing can be considered as an alternative to using histograms for approximating PDFs. However, it avoids the ambiguity of binning and yields a smooth analytic curve. For 1-dimensional data, each sample data point is replaced by a smooth Gaussian kernel. To create a PDF, all the kernels can be summed and then normalised to integrate to unity. The width of these smoothing Gaussians is still to be decided, but methods exist for optimising this factor. Each kernel can even be given a different width, calculated as a function of a quick local density estimate, in order to produce adaptive smoothing.

In data with more than one dimension, each sample point is replaced by a multivariate kernel. To help overcome the difficulties associated with the leaking of probability density into the wings of many-dimensional kernels, we replace the Gaussian with a more compact *Epanechnikov* kernel (Epanechnikov 1969),

$$K(\delta x_i) = \begin{cases} \frac{3}{4\lambda_i}\left(1 - \left(\frac{\delta x_i}{\lambda_i}\right)^2\right) & \text{for } -\lambda_i < \delta x_i < \lambda_i \\ 0 & \text{elsewhere,} \end{cases} \quad (3)$$

where we have reformatted the shapelet coefficients of each HDF galaxy into a data vector x_i, and δx_i are deviations in shapelet space from these real data points. In each case, i is a coefficient index running from 1 to 137. λ_i are smoothing widths which will be determined for each direction of our PDF space in §4.2. Isodensity contours of this kernel are multivariate ellipses whose axes are aligned with those of the co-ordinate axes (see figure 2). In general, they could be allowed to point in any direction (Sain 1999), but we do not find this to be necessary.

We implement an adaptive smoothing of our PDF by reparametrizing a_{nm} as $\frac{a_{nm}}{a_{00}}$. Given a constant λ_i, this creates an effective smoothing kernel for each object of widths $\lambda_i' = a_{00}\lambda_i$. This functional form is useful because the

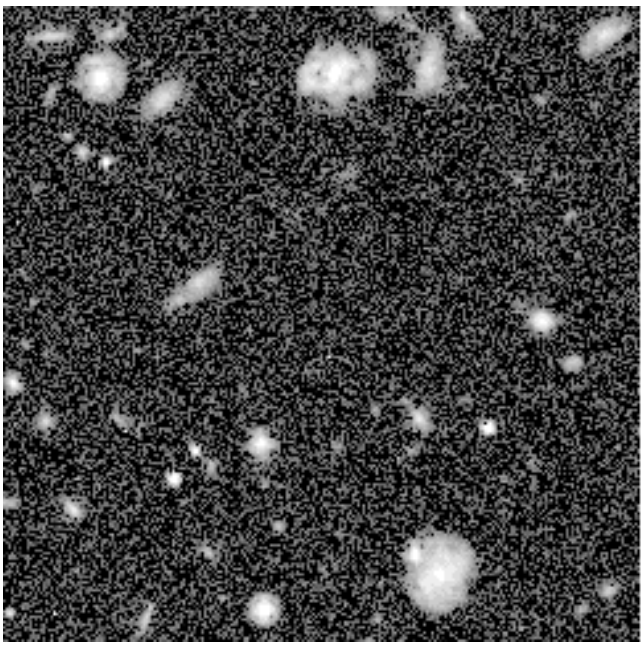

Figure 4. Sample HDF-depth simulated I-band image, 30"×30".

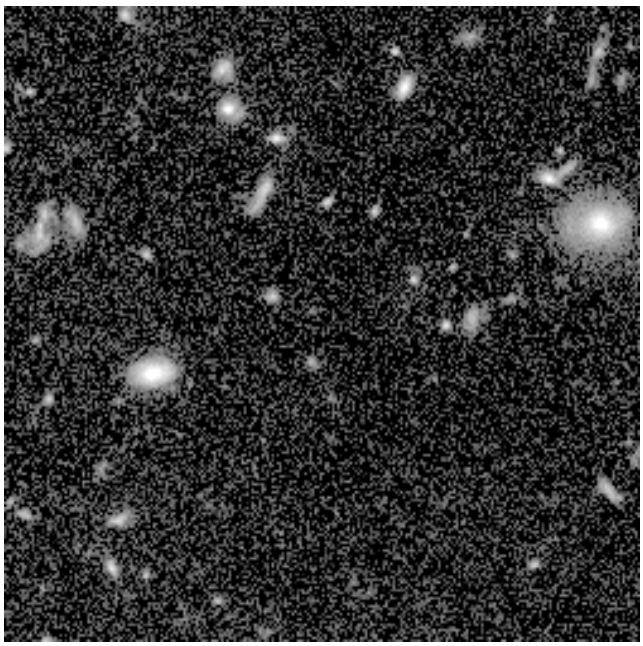

Figure 5. Section of the real HDF, with the same size and scale as figure 4.

brighter HDF objects are less frequent, and are therefore more isolated in probability space. Since a_{00} roughly correlates to total flux, we obtain a larger smoothing radius for brighter objects and better recover the underlying probability distribution. We will show that this recipe produces realistic morphologies in §5.

4.2 Monte-Carlo sampling

The advantage of the kernel smoothing approach is made clear by Silverman (1986), §6.4.1. Without resorting to costly numerical integration, he presents a quick bootstrap method to generate a Monte-Carlo sample from the PDF constructed from δ-functions and smoothed with kernels $K(\mathbf{x})$. The following steps were taken to smooth the parameter space of HDF galaxies:

$$\left.\begin{array}{ll} \textit{Step } 1: & \text{Randomly select one of the original HDF} \\ & \text{galaxies, uniformly and with replacement.} \\ \textit{Step } 2: & \text{Generate a small perturbation } \delta x_i \text{ from} \\ & \text{the probability density function } K(\delta x_i). \\ \textit{Step } 3: & \text{Add } \delta x_i \text{ to the shapelet coefficients } x_i \text{ of} \\ & \text{the HDF galaxy. This simulates a new} \\ & \text{galaxy, sampled from the overall PDF.} \end{array}\right\} \quad (4)$$

This approach is simply arrived at by regarding the PDF as a sum of small kernels, rather than one overall function. Individually, these are quick to compute. The dimensionality of the PDF can even be lowered for objects using fewer coefficients. Iterating the above procedure to generate sufficient objects for a simulated Hubble Deep Field takes only a few minutes on a 1GHz PC.

We must now consider how to choose the overall smoothing length λ_i. If $\lambda_i \equiv 0$, the kernel is a δ-function and the original HDF objects are recovered exactly. As $\lambda_i \to \infty$, the coefficients for simulated galaxies become random and the objects become unrealistic. In this limit, since no positive-definite constraint is ever imposed in the shapelet simulation, we find that simulated objects exhibit holes of negative flux. Figure 3 shows realisations of how a typical galaxy from the HDF is altered by increasingly large perturbations to its shapelet coefficients, showing negative flux for large λ_i perturbations.

We therefore require a choice of λ_i which is sufficiently large to produce new galaxies, yet sufficiently small to result in realistic morphological properties. By measuring the minimum flux of many different galaxy realisations, we find that in order to avoid negative flux, we require $\lambda_{phase} \lesssim 15^\circ$ and $\lambda_{moduli} \lesssim 4\times$[mean separation between nearest neighbours in that dimension]; with λ_i larger than this, the objects quickly acquire negative holes. For the purposes of this paper, we therefore fix λ_i to these limiting values, demonstrating in section 5 that this produces morphologically satisfactory results.

4.3 Image generation

A Monte Carlo population of simulated galaxies are extracted from the smoothed PDF of galaxy morphology according to procedure (4). Represented as shapelets, these are effectively stored as analytic functions. They can be convolved with a PSF that has also been decomposed into shapelets, using the matrix operation in paper II §3.1. The galaxies are then pixellated into an image with the same object number density (~ 320 per arcmin^2 at $I < 29$) and resolution (0.0398"/pixel) as the `DRIZZLE`d Hubble Deep Field. Galaxy rotations and centroid positions are randomly chosen. The correct size-magnitude distribution is automatically generated, as size and magnitude are two variables in the PDF (see the bottom-left panel of figure 2). Our images have a somewhat larger solid angle than the HDFs because the missing quarter from the WFPC 'L' is restored.

Stars can be added, given a magnitude distribution, by

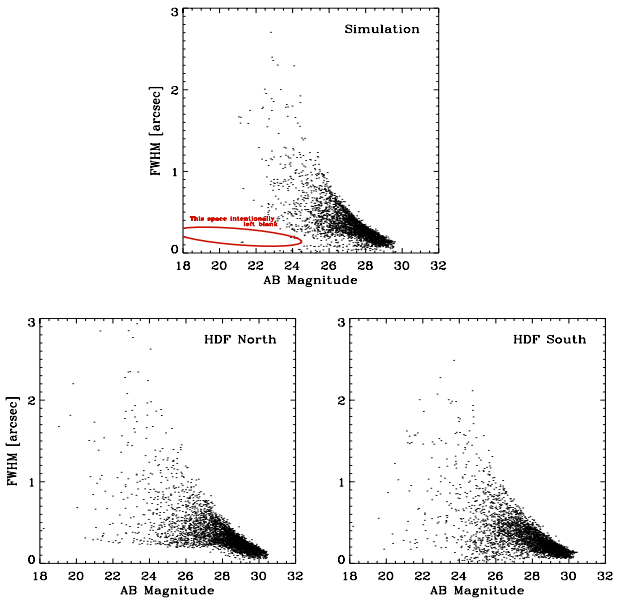

Figure 6. Size *vs* magnitude plane, measured with `SExtractor`. Top-left panel is for a simulated 7.5 square arcminute I-band image to HDF depth; the bottom panels show the HDFs themselves, calculated using the same `SExtractor` input parameters for reference. The stellar locus is omitted from all panels.

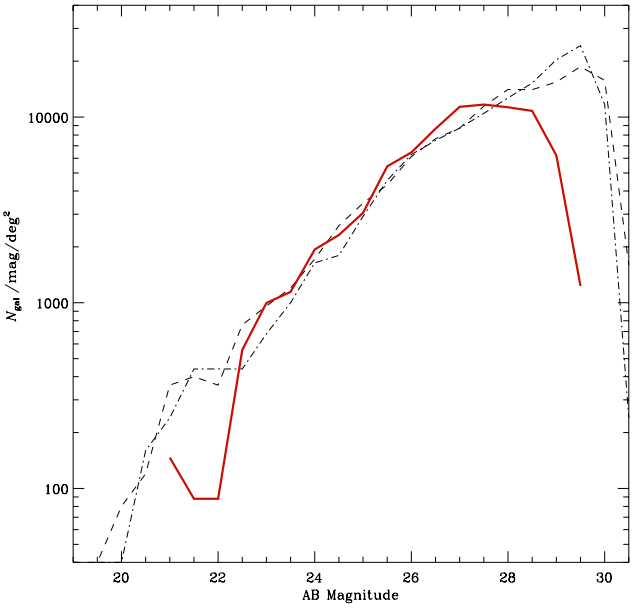

Figure 7. Number counts in simulated images (solid red), normalised by area on the sky. Also shown are number counts for the Hubble Deep Field North (dot-dashed) and South (dashed).

laying down the shapelet reconstruction of the PSF at the appropriate flux amplitude. Noise can also be added, in an amount appropriate to the desired exposure time. We have simply added photon counting noise (proportional to the square root of the raw pixel values), and Gaussian background noise (with amplitude determined from the HDF itself). However, it would be easy to add a background level, random cosmic rays and even instrumental distortions, the shearing for which could be performed conveniently in shapelet space before pixellisation. Final output is as a `fits` image. A small sample of a final image is displayed in figure 4. Larger images may be downloaded by anonymous `ftp` via `http://www.ast.cam.ac.uk/~rjm/shapelets`.

We note the wide range of galaxy morphologies and behaviours present in figure 4. In particular, features such as spiral arms, dust lanes and resolved knots of star formation are present, together with various radial profile shapes. The simulated galaxies look very similar to those in a similarly-scaled section of the HDF itself, reproduced in figure 5. We will examine quantitatively whether our simulation effectively mimics the morphology distribution of the HDF in §5.

It is important to consider the question of incompleteness for faint galaxies in our simulations. The total galaxy number density is equal to that detected in the HDF by construction. However, a discrepancy could nevertheless arise through either non-detections of faint HDF galaxies by `SExtractor` or non-convergence of the shapelet decompositions. We have chosen `SExtractor` parameters specifically to minimize the first effect (see §2.1) and the second is shown in §2.2 to be under control. However, the number counts of galaxies at the very faint end ($I \gtrsim 29$) are highly sensitive to the precise background noise properties. For example, the `DRIZZLE` algorithm recovers image resolution but aliases the noise. This can make it possible to detect slightly fainter objects and, because of the steep slope of the number counts at this depth, we choose not to consider galaxies below $I = 29$.

At the bright end, we also expect the simulations to be incomplete, since the HDFs were intentionally chosen as areas containing few large, bright galaxies. In the future, we will extend our simulations in this respect by incorporating 'Groth survey strip' (Groth *et al.* 1994) and ACS galaxies into the object source catalogue. One can also compensate for any known incompleteness by preferentially selecting for underrepresented galaxy types in procedure (4).

5 STATISTICAL TESTS AND RESULTS

We now demonstrate quantitatively that our simulated images are realistic, in the sense that commonly used morphology measures for our galaxies match those for galaxies in the HDFs. First, we consider the number counts and size distributions. For this purpose, we measure photometry and size measures using `SExtractor` (Bertin & Arnouts 1996). We also compare more detailed morphology measures, such as concentration (Bershady *et al.* 2000), asymmetry (Conselice *et al.* 2000a), clumpiness (Conselice *et al.* in preparation) and ellipticity. These are not expected to match by construction alone, as our shapelet measures do not directly represent these quantities. Thus, a comparison between these properties for simulated and real data is a measure of the realism of our simulations.

5.1 Size and magnitude

In order to carry out these tests, we first applied the `SExtractor` object-finding and shape measurement package on the version 2 reductions of the HDF-N and HDF-

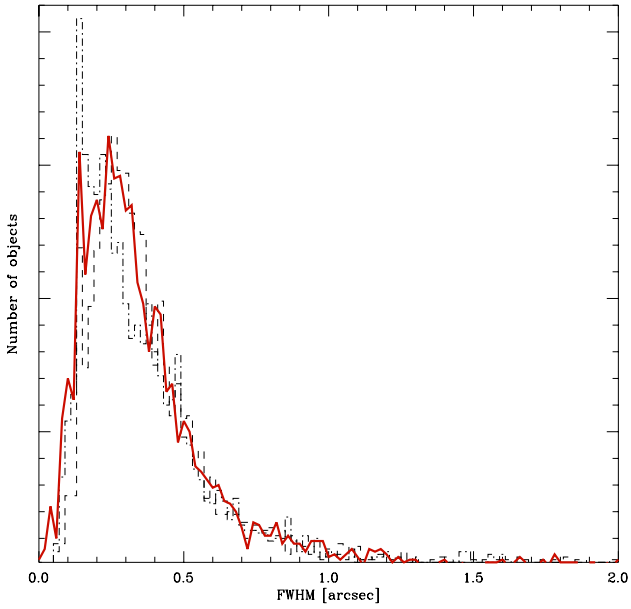

Figure 8. Size distribution of objects in a 6 square degree square simulated image with limiting magnitude $I < 29$ (solid red). Also shown are size distributions for the Hubble Deep Field North (dot-dashed) and South (dashed).

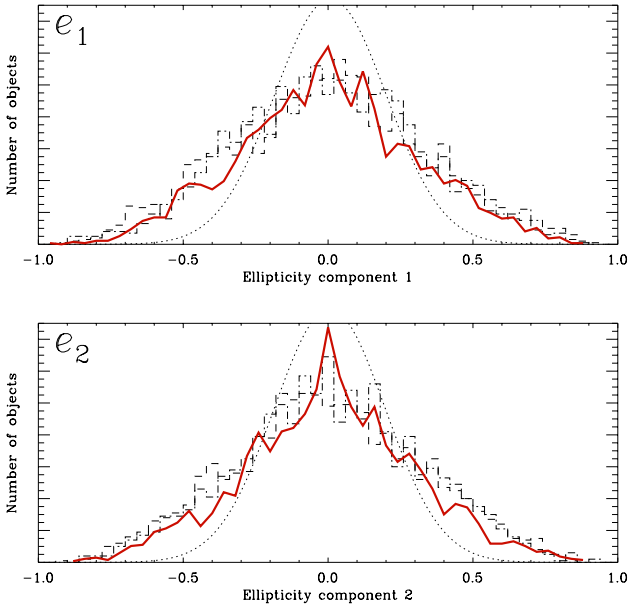

Figure 9. Ellipticity distribution, as defined in equations (5) and (6), of objects in a simulated 6 square arcminute I-band image with limiting magnitude $I < 29$ (solid line). Also shown is the ellipticity distribution for the Hubble Deep Field North (dot-dashed) & South (dashed), and a Gaussian with the same mean and rms (dotted).

S, together with a 7.5 square arcminute simulated image of the same depth. In each case, approximately 320 galaxies brighter than $I = 29$ were detected per arcmin^2. For all the galaxies, we extracted magnitudes (MAG_BEST) and sizes (FWHM_IMAGE).

Figure 6 compares the size *vs* magnitude distributions of the simulated image with those of the two HDFs, excluding the stars. Figure 7 shows more clearly the galaxy number counts in the three images, which match well over six or more orders of magnitude. The noise in this simulated image is not aliased in the same way as the `DRIZZLE` algorithm has acted upon real data (see §4.3). The number counts beyond $I \sim 29$ are highly sensitive to background noise properties, and are increased if we correlate the noise. Clearly `DRIZZLE` is something that needs further attention in future simulations.

For the present purposes, we look at brighter objects which are unaffected by such minor changes and apply magnitude cuts in subsequent comparisons. These cuts are at levels determined by the stability of an individual diagnostic to noise. Figure 8 compares the size distribution of the simulated objects brighter than $I < 29$ with that of the HDF galaxies, as found by `SExtractor`. We find that there is excellent agreement in the shape of this distribution. The median and standard deviation FWHM for galaxies in the simulation are 0.31" and 0.23". For the HDF-N with the same cuts, these figures are 0.32" and 0.24". For the HDF-S, we find 0.29" and 0.25".

5.2 Galaxy morphology diagnostics

The ellipticity of all the galaxies was also obtained from `SExtractor`. Following a convention in weak lensing literature, we here define the two components of ellipticity as

$$e_1 \equiv \frac{\text{A\_IMAGE}^2 - \text{B\_IMAGE}^2}{\text{A\_IMAGE}^2 + \text{B\_IMAGE}^2} \cos(2\times\text{THETA\_IMAGE}) \quad (5)$$

$$e_2 \equiv \frac{\text{A\_IMAGE}^2 - \text{B\_IMAGE}^2}{\text{A\_IMAGE}^2 + \text{B\_IMAGE}^2} \sin(2\times\text{THETA\_IMAGE}) \quad (6)$$

where A_IMAGE and B_IMAGE are the lengths of the major and minor axes of the ellipse, and THETA_IMAGE is the angle between the major axis and the horizontal. Figure 9 compares the ellipticity distribution of the real and simulated objects. Again, these are in excellent agreement, with a standard deviation in $e = \sqrt{e_1^2 + e_2^2}$ of 0.62 in the case of our simulated images, and 0.64 in the case of the real HDFs. For galaxies brighter than $I < 28$, the distributions are indistinguishable.

The simulated images were then passed through the morphology software of Concelice *et al.* (2002a), Bershady *et al.* (2000 and Conselice *et al.* (2003, submitted) in order to measure their concentration, asymmetry and clumpiness. We will first describe how these three quantities are calculated, and then compare the values obtained for these measures for real and simulated data.

The concentration index, C, is defined in terms of the ratio of the radii containing 80% (r_{80}) and 20% (r_{20}) of the object's total flux

$$C \equiv 5 \times \log\left(\frac{r_{80}}{r_{20}}\right) \ . \quad (7)$$

For the total flux, we use the flux within an aperture of size 1.5 times the Petrosian radius at $\eta = 0.2$ (Bershady *et al.* 2000). η is defined as the ratio of the surface brightness at a radius divided by the surface brightness integrated within the radius, such that at the centre of a galaxy, $\eta = 1$ and at

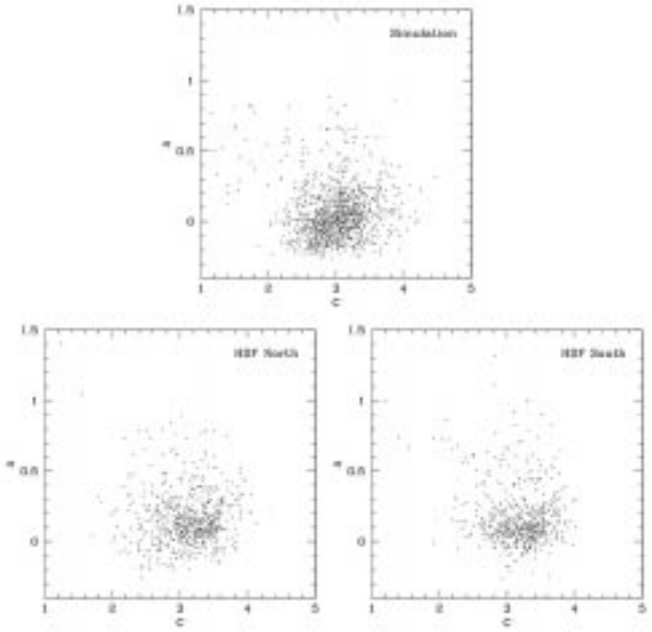

Figure 10. Top panel: concentration (C) *vs* asymmetry (A), as defined in equations (7) and (8), for a simulated 7.5 square arcminute I-band image with limiting magnitude $I = 26$. The two lower panels display C *vs* A for the real Hubble Deep Field North and South also down to $I = 26$.

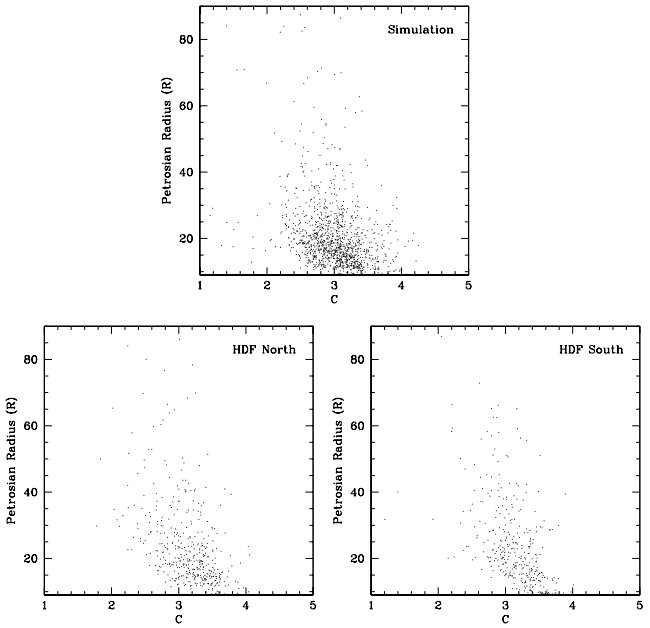

Figure 11. Top panel: Concentration (C) *vs* Petrosian radius (R) for a simulated 7.5 square arcminute I-band image. Bottom panels: C *vs* R for the Hubble Deep Field North and South. As in Figure 10, only galaxies with $I < 26$ are plotted.

the very edge of a galaxy (where its surface brightness is 0), $\eta = 0$.

The quantitative values of C for real galaxies range from approximately 2 to 5; galaxies with $C > 4$ are usually ellipticals or spheroidal systems. A galaxy with an $r^{1/4}$ profile has $C = 3.2$, and a purely exponential disc galaxy has $C = 2.5$ (Conselice *et al.* 2002). Objects less concentrated than this are shown by Graham *et al.* (2001) to be those with low central surface brightnesses and internal velocity dispersions. Low concentration values are also found for dwarf galaxies (*e.g.* Conselice *et al.* 2002). The concentration index thus correlates, within some scatter, with the total mass of a galaxy.

The asymmetry index used in this paper, A_{180} as defined by Conselice *et al.* (2000a), is produced by rotating an image by 180° and subtracting the result from the original image; the residuals are then quantified by

$$A \equiv \min\left[\frac{\sum |I_{x,y} - I^{180}_{x,y}|}{\sum |I_{x,y}|}\right] - \min\left[\frac{\sum |B_{x,y} - B^{180}_{x,y}|}{\sum |I_{x,y}|}\right]. \quad (8)$$

where $I_{x,y}$ is the galaxy surface brightness in the (x, y) pixel in the image, $B_{x,y}$ the sky background in the same pixel, and superscripts denote rotations. Sums are over all pixels and minimisation is over different choices of the centre of rotation.

A is sensitive to any feature in a galaxy that produces asymmetries in light distributions, such as star-formation, galaxy interactions/mergers, and projection effects such as dust lanes. There is a general correlation between the asymmetry value and the (B-V) colour (Conselice *et al.* 2000a). Since most galaxies are not edge-on systems, star formation and galaxy interactions/mergers are the dominant effects that produce asymmetries in real galaxies. These two effects

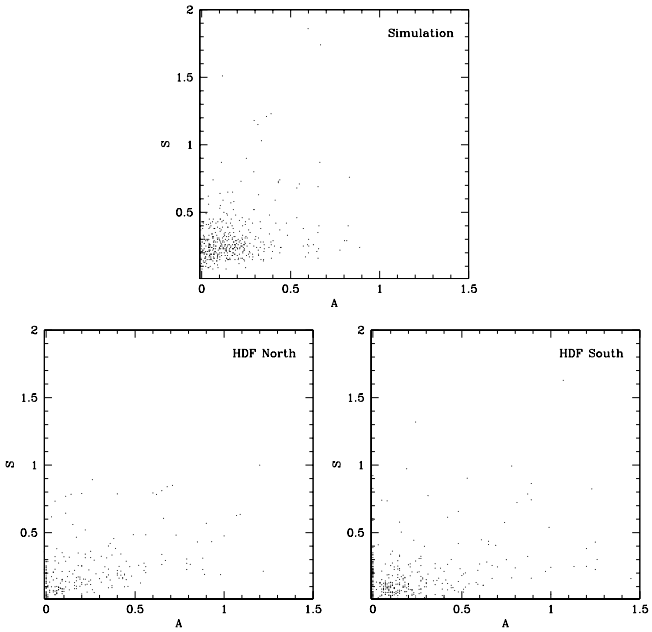

Figure 12. Top panel: Clumpiness (S) *vs* Asymmetry (A), as defined in equations (9) and (8), for a simulated 7.5 square arcminute I-band image. Bottom panels: S *vs* A for the Hubble Deep Field North and South. Only galaxies brighter than $I = 25$ are plotted.

can often be distinguished, however. Systems with asymmetries $A > 0.35$ are generally found to be due to interactions or mergers (Conselice *et al.* 2000b). However, other merger events can have more modest asymmetry values. From this and more detailed studies of the asymmetry index, it has been concluded that A is most sensitive to bulk structures in galaxies (Conselice, in preparation).

The clumpiness parameter, S, is a measure of the high-spatial frequency component of galaxies. It is sensitive to star formation, as it correlates very well with Hα equivalent widths; it also correlates to a lesser degree with broad-band colors (Conselice, in preparation). The clumpiness is computed by smoothing a galaxy's image with a smoothing length σ and then subtracting this smoothed image $I^{\sigma}_{x,y}$ from the original. This leaves a residual map with only the high-spatial features left. After correcting for the background, $\mathrm{B_{x,y}}$, the clumpiness, S, is computed as:

$$S \equiv \sum_{xy} \left\{ \frac{[I_{x,y} - I^{\sigma}_{x,y}] - B_{x,y}}{I_{x,y}} \right\} . \quad (9)$$

We also use the Petrosian radius R (Petrosian 1976) to characterize the galaxies, with parameter $\eta = 0.2$. The Petrosian radius is found to be a better index than the `SExtractor` FWHM radius for determining morphological sizes, as `SExtractor` radii are based on isophotal thresholds which will represent different physical distances from the galactic centre depending on the distance to the galaxy. Because η is a ratio of surface brightnesses in a given galaxy, the run of η with r in a galaxy is immune to many such types of systematic effects (Sandage & Perelmuter 1990) and Petrosian radii are found to be a stable method for deriving morphological parameters independent of distance (Bershady *et al.* 2000).

We are now in a position to compare the measurements for C, A, S and R for real and simulated images. Projections from this morphological parameter space for simulated and real data points are displayed in Figures 10–12.

As can be seen from the scatter in the plots, the agreement between simulation and reality is rather good. For example, the mean concentration C of our simulated galaxies is found to be 3.00, while for real galaxies it is 3.11. The standard deviation of the concentration index is found to be 0.52 in the case of simulations and 0.45 in the case of real data. A criticism one might have been tempted to level at shapelets is that they are a truncated series, which may not stretch far enough spatially to capture the extended wings of objects. Clearly our algorithm sets $n_{\max}$ high enough to avoid this problem while still modelling the HDF galaxies using just a few coefficients.

Turning now to the asymmetry parameter A, we find that the simulated galaxies have mean A =0.06 and standard deviation 0.21 while the real galaxies have mean 0.17 and standard deviation 0.20. We note that the population of simulated galaxies have marginally lower asymmetries than the real data, although within 1σ. This may be an artifact of having a finite $n_{\max}$ cutoff in the shapelet modelling, or may be due to the fact that the positions of the simulated galaxies are uncorrelated. Hence there are no simulated merger events except those that are sufficiently well advanced to be counted as one object by `SExtractor`.

We now consider the clumpiness measure, S. We find that the simulated galaxies have a mean clumpiness of 0.27 with standard deviation 0.21, whereas the real galaxies have mean 0.23 and standard deviation 0.37. Thus while the agreement is good, the simulated galaxies appear slightly too clumpy. We suspect that this is simply due to the fact that the noise in adjacent simulated pixels is not correlated. The HDFs themselves have been `DRIZZLE`d in order to achieve this high resolution, a process which also aliases the image. As a simple approximation, we have tried smoothing the noise slightly, by a top hat kernel 3 pixels wide. This process does indeed remove the slight disparity in clumpiness, but creates many false detections of faint circular objects on the limit of detection.

Thus we conclude that our shapelet simulations obtain very similar morphology distributions to those found in real data. This is most encouraging as these were not arranged by construction, and the level of realism seen here is a vindication of both the shapelet modelling of galaxies and our perturbation of parameters in shapelet space. We can therefore use shapelets as a tool for investigating galaxy morphology and creating simulated sets of realistic imaging data.

5.3 Comparison to other methods

There have been many packages in the literature which simulate astronomical observations, including `Skymaker` (see Erben *et al.* 2001) and `artdata` in `IRAF` (Tody 1993). These typically parametrize galaxy shapes using simple physical models such as ellipses with de Vaucouleur or exponential profiles. The smooth variation allowed for these parameters enables them to generate an unlimited number of galaxies to simulate images from ground-based telescopes. Unfortunately, deep images from HST contain galaxies with resolved features more complex that their elliptical models, and such simulations are useful in only a limited regime.

This was realised by Bouwens, Broadhurst & Silk (1998), who designed simulations to investigate the evolution of galaxy morphology in the HDF. Indeed, their work succeeds in ruling out pure luminosity evolution of galaxies: which precisely demonstrates the need for deep image simulations to contain more irregular and asymmetric morphologies. Their method repeatedly places the few brightest HDF galaxies onto a simulated image, and is similar to that which ours would have been, had we left the PDF as an (unsmoothed) sum of δ-functions. Some physics can be added to rescale and redshift these few sources, but it remains a very small population if we want to simulate a large survey. Creating realistic images was not the intention of Bouwens, Broadhurst & Silk (1998) and, in this framework, would require the addition of more physics (*e.g.* galaxy evolution, star formation histories, redshift distributions, *etc.*). This is not appropriate for our present objectives.

Our technique attempts to capture the best aspects of both methods by defining a smooth parameter space which will yield many unique galaxies, but which contains a rich diversity of morphologies (potentially any morphology in fact, since the shapelet basis functions are complete). Since the parameter space is populated by statistical rather than physical arguments, it is the many tests to which we have subjected our simulated images in §5 that demonstrate the validity of our method. We find a regime spanning six or-

ders of magnitude in brightness where our simulations are valid and the statistical properties match well those of real data. The ability to produce simulated galaxies with realistic morphologies is a significant advance.

6 CONCLUSIONS

We have presented a method of generating simulated deep sky images for an arbitrarily large survey area, as might be observed with extended observations with the Hubble Space Telescope. These simulated images are populated with all morphological types of galaxies, based upon those in the Hubble Deep Fields (HDFs).

The simulated galaxies are drawn from a multi-variate distribution of realistic morphologies, described using the shapelet formalism (Refregier 2003). In order to generate this morphology distribution, we decompose all HDF galaxies into shapelet components using least-squares fitting. We optimise this decomposition by finding scale length β and number of modes $n_{\max}$ which produces a best shapelet coefficient fit. The resulting coefficients of HDF galaxies form a cloud of points in shapelet space. These are replaced by smooth kernels in order to recover the underlying probability distribution of real galaxy morphologies. The smooth distribution is resampled, using an unbiased Monte-Carlo technique, to obtain new galaxies.

We place the simulated galaxies onto HDF size images, including effects such as PSF, pixellisation, poisson noise and Gaussian background noise. The level of detail in the resulting simulated galaxies includes features such as realistic radial profiles, spiral arms, dust lanes and resolved knots of star formation.

The global morphological properties of the simulated galaxy population must match those of real galaxies if our simulations are to be useful. We have demonstrated that this is the case by comparing various properties of simulated and real galaxies, including number counts with magnitude, size distribution, ellipticity distribution, and morphological diagnostic tools such as concentration, asymmetry and clumpiness indices.

An important application for the resulting simulated images is presented in Massey *et al.* (2003b), where the images are used to calibrate the sensitivity to weak gravitational lensing of the proposed SNAP satellite. However, the simulations presented here are in no way specific to gravitational lensing, and may be used for testing image analysis in various branches of astronomy. Further simulated images and catalogues are available from the authors.

A useful extension to this work will be to include 'Groth survey strip' (Groth *et al.* 1994) galaxies and ACS data in constructing the morphology probability distribution. This will add to future simulations a more extensive sample of large, bright galaxies, improving the fidelity of the simulations in this region of parameter space. A method is also in development to generate multi-colour simulated images using several HDF passbands. A by-product of this work is a complete morphological catalogue of all the HDF galaxies in shapelet space. This catalogue will be used in a future paper concerning the automated morphological classification of galaxies at high redshift.

ACKNOWLEDGEMENTS

We thank the Raymond and Beverly Sackler fund for travel support. AR was supported in Cambridge by a PPARC advanced fellowship. DJB was supported by a PPARC fellowship. The authors would like to thank Tzu-Ching Chang for her help in developing the shapelets method and code. This would run much slower without Sarah Bridle and Phil Marshall's insightful statistical trickery. Thanks to Jason Rhodes for help implementing the morphological tests. We are also grateful to Richard Ellis, Josh Frieman, Andy Fruchter, Eric Gawiser and Jean-Paul Kneib for ideas, comments and enthusiasm throughout this work.

REFERENCES

Abraham R. *et al.* 1996, ApJS, 107, 1
Bacon D., Refregier A., Clowe D. & Ellis R. 2001, MNRAS, 325, 1065
Bartelmann M. & Schneider P. 2001, Physics Reports, 340, 291
Bernstein G. & Jarvis M. 2002, AJ, 123, 583
Bershady M., Jangren A. & Conselice C. 2000, AJ, 119, 2645
Bertin E. & Arnouts S. 1996, A&AS, 117, 393
Blinnikov S. & Moessner R. 1998, A&AS, 130, 193
Boas M. 1983, "Mathematical Methods in the Physical Sciences", Wiley.
Bouwens R., Broadhurst T. & Silk J. 1998a, ApJ, 506, 557
Bouwens R., Broadhurst T. & Silk J. 1998b, ApJ, 506, 579
Conselice C., Bershady M. & Jangren A. 2000a, ApJ, 529, 886
Conselice C., Bershady M. & Gallagher J. 2000b, A&A, 354, 21L
Conselice C., Gallagher J. & Wyse R. 2002, AJ, 123, 2246
Driver S. 1999 in Looking Deep in the Southern Sky, eds F. Morganti and W. Couch. Berlin: Springer-Verlag p.280
Epanechnikov, V. 1969, "Nonparametric Estimation of a Multivariate Probability Density", Theory of Probability and Its Applications, 14, 153
Erben T., van Waerbeke L., Bertin E., Mellier Y. & Schneider P. 2001, A&A, 366, 717
Francis P. & Wills B. 1999, in "Quasars and Cosmology", A.S.P. Conference Series 1999, eds. G. Ferland, J. Baldwin,(San Francisco: ASP)
Fruchter A. & Hook R. 2002, PASP, 114, 144
Gerhard O. 1993, MNRAS, 265, 213
Groth E. *et al.* 1994, BAAS, 185, 5309
Hubble E. 1926, ApJ, 64, 321
Kaiser N., Squires G. & Broadhurst T. 1995, ApJ, 449, 460
Krist J. 1995, ADASS IV, A.S.P. Conference Series, Vol. 349
Krist J., & Hook R. 1997, The Tiny Tim User's Guide (Baltimore: STScI)
Lupton R. 1993, Statistics in Theory and Practice (Princeton U. Press: Princeton)
Marleau F. & Simard L. 1998, ApJ, 507, 585
Marshall P. 2003, in preparation
Massey R. *et al.* 2003b, ApJ submitted
Mellier Y. 1999, ARA&A, 37, 127
Meyer Y. 1993, "Wavelets: Algorithms and Applications", Society for Industrial and Applied Mathematics, Philadelphia
Myller-Lebedeff. W. 1908, Mathematische Annalen, 64, 388
Odewahn S., Windhorst R., Driver S. & Keel W. 1996, ApJL, 472, 13
Peng C., Ho L., Impey C. & Rix H.-W. 2002, AJ in press.
Petrosian V. 1976, ApJ, 209, 1L
Piña R. & Puetter R. 1993, P.A.S.P., 105, 630
Ratnatunga K., Griffiths R. & Ostrander E. 1999, AJ, 118 86
Refregier A. 2003, MNRAS, 338, 35
Refregier A. & Bacon, D. 2003, MNRAS, 338, 48

Rhodes J., Refregier A. & Groth E. 2001, ApJ, 552, 85
Sandage A. 1961, "The Hubble Atlas of Galaxies" (Carnegie Inst. Washington Publ. 618) (Washington: Carnegie Inst.)
Sain S. 1999, "Multivariate Locally Adaptive Density Estimation", Technical Report, Department of Statistical Science, Southern Methodist University
Sérsic J. 1968, "Atlas de Galaxias Australes" (Cordoba: Obs. Astronomico)
Silverman B. 1986, "Density Estimation for Statistics" and Data Analysis (Chapman and Hall, London)
Simard L. 1998, ADASS VII, A.S.P. Conference Series, Vol. 145, 108
Tody D. 1993, "IRAF in the Nineties" in ADASS II, A.S.P. Conference Series, Vol 52, eds. R. Hanisch, R. Brissenden, & J. Barnes, 173
Van der Marel R. & Franx M. 1993, ApJ, 407, 525
Van der Marel R. *et al.* 1994, MNRAS, 268, 521
de Vaucouleurs G. 1959, Hand. Physik, 53, 275
Williams R. *et al.* 1996, AJ 112, 1335
Williams R. *et al.* 1998, A&AS, 193, 7501
York D. 2000, AJ, 120, 1579

This paper has been produced using the Royal Astronomical Society/Blackwell Science LaTeX style file.